HIGH DENSITY STORAGE SCHEME FOR SEMICONDUCTOR MEMORY

Technical Field of the Invention

The present invention relates generally to semiconductor memory and in particular to data compression and decompression in a dynamic random access memory device.

Background of the Invention

Semiconductor memory devices such as dynamic random access memory (DRAM) devices are widely used to store data in computers and electronic products. One important criterion in a DRAM device is storage density. As semiconductor technology advances, designers strive to design new generation of DRAM device with a higher storage density.

There are many methods used in designing a DRAM device to achieve a higher density. One method is reducing the size of each of millions of cells in a DRAM device thus increasing the number of cells and consequently increasing the overall storage density of the device. Another method is stacking the cells vertically, this in effect, doubles the storage capacity. There are other methods of designing a DRAM device to achieve high storage density.

There is a need for a method to achieve even higher storage density for a DRAM device using other techniques.

Summary of the Invention

The present invention increases a storage density of a semiconductor memory device. In particular, the invention includes an apparatus and method of compressing and decompressing data in a DRAM device.

In one embodiment of the invention, the memory device includes a main memory, an input/output buffer, a cache memory connected between the main memory

and the input/output buffer, and a compression and decompression engine connected between the main memory and the cache memory.

In another embodiment of the invention, the memory device includes a main memory, a cache memory connected to the main memory, a compression and decompression engine connected to the main memory and the cache memory, and an error detection and correction engine connected to the main memory and the compression and decompression engine.

Yet another embodiment of the invention provides a method of increasing a storage density of a memory device. The method comprises forming a main memory in a semiconductor chip, forming a cache memory, forming a compression and decompression engine in the same chip, and connecting the compression and decompression engine between the main memory and the cache memory.

Brief Description of the Drawings

Figure 1 illustrates a prior art a memory device;

Figure 2 illustrates a memory device of the present invention.

Figure 3 illustrates a memory device of Figure 2 including an error detection and correction (ECC) engine;

Figure 4 illustrates another embodiment of a memory device.

Figure 5 illustrates a computer system incorporating the memory devices of Figures 2-4; and

Figure 6 illustrates an exemplary environment including memory device of the present invention.

Detailed Description of the Invention

The following detailed description of the preferred embodiments refers to the accompanying drawings which form a part hereof, and shows by way of illustration specific preferred embodiments in which the inventions may be practiced. These embodiments are described in sufficient detail to enable those skilled in the art to practice the invention, and it is to be understood that other embodiments may be utilized

and that logical, mechanical and electrical changes may be made without departing from the spirit and scope of the present invention. The following detailed description is, therefore, not to be taken in a limiting sense, and the scope of the invention is defined only by the appended claims.

Figure 1 illustrates a prior art a memory device 100. Memory device 100 includes a cache memory 102 connected between a main memory 104, and an input/output (I/O) buffer 106. Main memory 104 typically comprises dynamic random access memory (DRAM) devices and comprises one or more memory banks, indicated by BANK 1-N. Each of the memory banks BANK 1-N comprises a plurality of memory cells arranged in rows and columns. Row decode circuit 108 and column decode circuit 110 access the rows and columns in response to an address, provided on address bus (ADDRESS) by an external controller, such as a microprocessor. I/O data buffers 106 connects to data communication lines (DATA) for bi-directional data communication with main memory 104. A memory controller 112 controls data communication between the memory 100 and external devices by responding to control signals (CONTROL SIGNALS).

Cache memory 102 comprises a plurality of fast static registers or channels, such as channels 102a-n. Each of the channels 102a-n is controlled by a channel controller, indicated by 103a-n. Because each of the channels 102a-n has its own controller , the channel 102a-n operates independently from each other and provide fast access paths between main memory 104 and I/O buffer 106. The independent operation and fast access path of each of the channels collectively result in a higher memory bus bandwidth and reduced data access latency of the memory device. A memory device which includes a cache memory such as that of memory device 100 is commonly referred to as virtual channel memory (VCM) device.

Figure 2 illustrates a memory device 200 of the invention. Memory device 200 comprises a memory chip 201. Memory device is similar to memory device 100 of Figure 1 with an exception of an additional feature indicated by a compression and decompression (C/D) engine 214. C/D engine 214 itself is well known to those skilled in the art. C/D engine 214 is connected between a main memory 204 and a cache

memory 202 which is connected to an I/O buffer 206. Main memory comprises one or more addressable memory banks, indicated by BANK 1-N. Each of the memory banks BANK 1-N comprises a plurality of memory cells arranged in rows and columns. Row decode circuit 208 and column decode circuit 210 access the rows and columns in response to an address, provided on address bus (ADDRESS) by an external controller, such as a microprocessor. I/O data buffers 206 connects to data communication lines (DATA) for bi-directional data communication with main memory 204. In addition, a memory controller 212 controls data communication between the memory 200 and external devices by responding to control signals (CONTROL SIGNALS).

Cache memory 202 may be selected from any type of memory but usually a static random access memory (SRAM) and normally operates at a faster speed than main memory 204, which is typical a dynamic random access memory (DRAM). Cache memory 202 may comprise one or more registers or channels as indicated in the Figures as channels 206a-n. Each of the channels is controlled by a channel controller 203a-n.

The inclusion of C/D engine 214 in a single chip, chip 201, with main memory 204 and cache memory 202 is practicable in this memory device because cache memory 204, having reduced data access latency, would hide or compensate any data access latency associated with C/D engine 214. Furthermore, the inclusion of C/D engine on the same chip with the main memory and the cache memory increases the operational speed of the memory device by eliminating off-chip drivers and connections.

As its name indicates, a C/D engine is a device which compresses and decompresses data using a hardware encoding scheme such as a Lempel Ziv encoding or other industry standard encoding schemes. One advantage of using a C/D engine, such as C/D engine 214, in a DRAM device is that data is compressed through the C/D before it enters main memory 204. This in effect increases a storage density of main memory 214.

In operation, I/O buffer 206 receives data from data communication lines DATA. Cache memory 202 processes the data and produces processed data. C/D engine 214 receives the processed data from cache memory 202 and compresses the

data before it is stored or written into main memory 204. The data stored in main memory 204 can be read and decompressed.

Figure 3 illustrates another memory device according to the invention. Memory device 300 is similar to memory 200 of Figure 2 with the exception of an error detection and correction (ECC) engine 302. ECC engine 302 is connected between C/D engine 214 and main memory 204. As in the case with the C/D engine, several types of ECC engines are well known to those skilled in the art. ECC engine 302, in memory device 300 will ensure the accuracy of data without substantially compromising a functional operation of the memory device, such as speed or latency, because cache memory 214 would reduce or compensate any latency which the ECC engine may produce. Furthermore, all the components of memory device 300 in Figure 3 are in a single chip, indicated by chip 301.

An ECC engine is a device which performs a process of detecting for error and correcting the error to ensure the accuracy and integrity data during a data transmission. Any ECC engine using Hamming Code, Reed-Solomon Code or other techniques can be used as ECC engine 302 in memory device 300 of the invention.

The inclusion of the C/D engine and the ECC engine in a memory device of the present invention as described above is not limited to memory devices having a cache memory or virtual channel memory. The C/D and the ECC engines may also be included in other memory devices including, but not limited to Double Data Rate synchronous DRAM (DDR SDRAM) or DRAM devices similar to RDRAM (TM) made by Rambus (TM) corporation.

Figure 4 illustrates a memory device of the invention having a C/D and the ECC engines included in a DRAM device other than a virtual channel memory device. Memory device 400 may be a DDR SDRAM device or an RDRAM (TM) device with the exception of C/D engine 414 and the ECC engine 402 connected between a main memory 404 and I/O buffer 406. Main memory comprises one or more addressable memory banks, indicated by BANK 1-N. Each of the memory banks BANK 1-N comprises a plurality of memory cells arranged in rows and columns. Row decode circuit 408 and column decode circuit 410 access the rows and columns in response to

an address, provided on address bus (ADDRESS) by an external controller, such as a microprocessor. I/O data buffers 406 connect to data communication lines (DATA) for bi-directional data communication with main memory 404. A memory controller 412 controls the operation of the memory via control signals, indicated as CONTROL SIGNALS. In addition, all the components of memory device 400 are in a single chip, indicated by chip 401.

Figure 5 illustrates a system 500 which includes a memory device 502 connected to an external processor 504. Memory device 502 may be one or more of any of the memory devices of the present invention described above. Memory 502 communicates with processor 504 via an address bus (ADDRESS) and a data bus (DATA) and control signals which include, but are not limited to, a Chip Select (CS*), a Clock (CLK), Row Access Strobe (RAS*), Column Access Strobe (CAS*), Write Enable (WE*), Clock Enable (CKE).

Figure 6 illustrates an embodiment of an exemplary environment of the invention. The embodiment includes a computer system 600 including one or more of the memory devices of the present invention. Computer 600 comprises a processor 602 connected to a memory device 604. Memory device 604 may be any of the memory devices of the invention described above. Memory device 604 is used for storing data or for other application within the system. Because of its high density, high bandwidth and reduced data access latency, memory device 604 can be used as random access memory of the computer system to improve the overall performance of the computer system. Memory device 604 can also improve the processing of graphic information or video information by storing more information and operating at reduced latency. Therefore memory device can be connected to a graphic control card 606 or a video control card 608 for use in processing graphic or video information.

Conclusion

Although specific embodiments have been illustrated and described herein, it will be appreciated by those of ordinary skill in the art that any arrangement which is calculated to achieve the same purpose may be substituted for the specific embodiment

shown. This application is intended to cover any adaptations or variations of the present invention. Therefore, it is intended that this invention be limited only by the claims and the equivalents thereof.